

# PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2001-195087

(43)Date of publication of application : 19.07.2001

(51)Int.Cl. G10L 15/28  
G06F 3/16  
G10L 15/14  
G10L 15/18  
G10L 15/00  
G10L 15/22

(21)Application number : 2000-000987

(71)Applicant : MITSUBISHI ELECTRIC CORP

(22)Date of filing : 06.01.2000

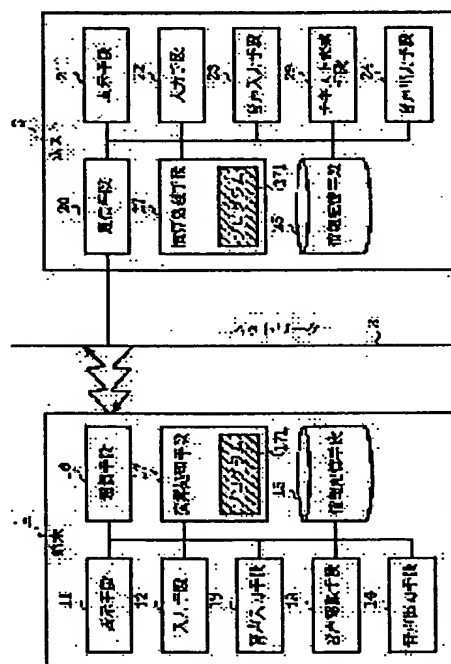
(72)Inventor : ABE YOSHIHARU

## (54) VOICE RECOGNITION SYSTEM

### (57)Abstract:

**PROBLEM TO BE SOLVED:** To obtain high recognition precision even when the load of a network is small and voice made up with a large vocabulary is inputted.

**SOLUTION:** A voice recognition means 18 of a terminal 1 analyzes voice signals from a voice input means 13 and extracts a syllable train in which a syllable is made into a basic unit. A communication means 16 transmits the syllable train to a host 2 through a network 3. A text converting means 29 of the host 2 inputs the syllable train received by a communication means 26, obtains a word train to match with the syllable train by using a difference model and a language model and outputs the word train as a text. The means 26 transmits back the text to the terminal 1 and a display means 11 of the terminal 1 displays the text.



## LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号  
特開2001-195087  
(P2001-195087A)

(43)公開日 平成13年7月19日(2001.7.19)

(51)Int.Cl. <sup>7</sup>	識別記号	F I	チーワード*(参考)
G 1 0 L 15/28		G 0 6 F 3/16	3 2 0 H 5 D 0 1 5
G 0 6 F 3/16	3 2 0	G 1 0 L 3/00	5 7 1 A
G 1 0 L 15/14			5 3 5 B
15/18			5 3 7 C
15/00			5 5 1 A

審査請求 未請求 請求項の数10 O L (全 7 頁) 最終頁に続く

(21)出願番号 特願2000-987(P2000-987)  
(22)出願日 平成12年1月6日(2000.1.6)

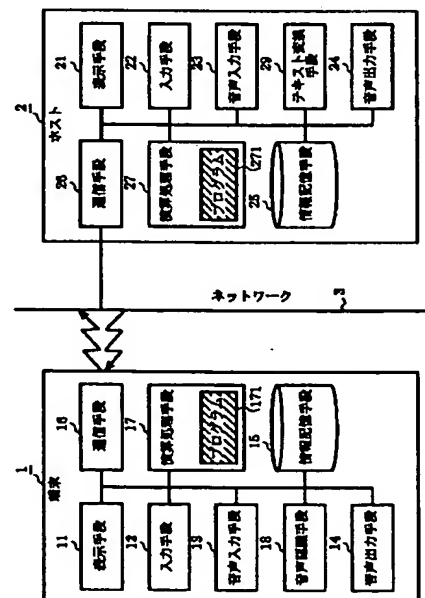
(71)出願人 000006013  
三菱電機株式会社  
東京都千代田区丸の内二丁目2番3号  
(72)発明者 阿部 芳春  
東京都千代田区丸の内二丁目2番3号 三  
菱電機株式会社内  
(74)代理人 100066474  
弁理士 田澤 博昭 (外1名)  
Fターム(参考) 5D015 B802 CC14 DD03 HH23 KK02  
LL04

(54)【発明の名称】 音声認識システム

(57)【要約】

【課題】 ネットワークの負荷が小さく、大語彙からなる入力音声の場合にも高い認識精度を得る。

【解決手段】 端末1の音声認識手段18は、音声入力手段13からの音声信号を分析し音節を基本単位とする音節列を抽出する。通信手段16は音節列をネットワーク3を介してホスト2に送信する。ホスト2のテキスト変換手段29は、通信手段26が受信した音節列を入力し、差分モデルと言語モデルを使用して、音節列に整合する単語列を求めてテキストとして出力する。通信手段26はテキストを端末1に送信し、端末1の表示手段11がテキストを表示する。



## 【特許請求の範囲】

【請求項1】 ネットワークを介して接続された端末とホストからなる音声認識システムにおいて、上記端末が、入力した音声信号から、類似の状態が共有化された音素HMM (Hidden Markov Model) を記憶した音響モデルを使用し、基本単位の列を抽出して上記ホストに送信し、上記ホストが、基本単位の列が示す誤認識の確率を記憶した差分モデルと、単語の連鎖の出現確率を記憶した言語モデルを使用して、受信した上記基本単位の列に整合する単語列を求め、そのテキストを上記端末に送信し、上記端末が受信したテキストを表示することを特徴とする音声認識システム。

【請求項2】 端末がユーザからの指示に基づき表示しているテキストを修正することを特徴とする請求項1記載の音声認識システム。

【請求項3】 端末が、表示しているテキストを他の端末又は他のホストに送信することを特徴とする請求項1又は請求項2記載の音声認識システム。

【請求項4】 基本単位として音素を使用することを特徴とする請求項1記載の音声認識システム。

【請求項5】 基本単位として音節を使用することを特徴とする請求項1記載の音声認識システム。

【請求項6】 基本単位として音素HMMの並びで表現されるサブワードを使用することを特徴とする請求項1記載の音声認識システム。

【請求項7】 端末が、ユーザからの音声を入力しデジタル化された音声信号を出力する音声入力手段と、上記音声信号から、類似の状態が共有化された音素HMMを記憶した音響モデルを使用して、基本単位の列を抽出する音声認識手段と、抽出された上記基本単位の列をホストに送信すると共に、上記ホストから送信されたテキストを受信する通信手段と、受信した上記テキストを表示する表示手段とを備えたことを特徴とする請求項1記載の音声認識システム。

【請求項8】 音声認識手段が、音声信号を入力して、特徴パラメータ時系列と短時間パワー時系列を抽出する音声分析手段と、上記短時間パワー時系列に基づき、音声パワーの存在する可能性の高い音声区間を検出して、上記音声区間内の特徴パラメータ時系列を出力する音声区間検出手段と、類似の状態が共有化された音素HMMを記憶した音響モデルと、この音響モデルに記憶されている音素HMMの組合せの中から、上記音声区間検出手段が出力した特徴パラメータ時系列に整合する基本単位の列を抽出するビタビデコーダとを備えたことを特徴とする請求項7記載の音声認識システム。

【請求項9】 ホストが、

端末から送信された基本単位の列を受信すると共に、求めたテキストを上記端末に送信する通信手段と、基本単位の列が示す誤認識の確率を記憶した差分モデルと、単語の連鎖の出現確率を記憶した言語モデルを使用して、受信した上記基本単位の列に整合する単語列を求め、テキストとして出力するテキスト変換手段とを備えたことを特徴とする請求項1記載の音声認識システム。

【請求項10】 テキスト変換手段が、

基本単位の列が示す誤認識の確率を記憶した差分モデルと、テキストを構成する単語の連鎖の出現確率を記憶した言語モデルと、

受信した基本単位の列を入力し、上記差分モデルが記憶した基本単位の列が示す誤認識の確率と、上記言語モデルが記憶したテキストを構成した単語の連鎖の出現確率に基づき、上記基本単位の列に整合する単語列を求め、テキストとして出力するスタックデコーダとを備えたことを特徴とする請求項9記載の音声認識システム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 この発明は、無線・有線に関わらずネットワークで接続された端末とホストからなる音声認識システムに関するものである。

【0002】

【従来の技術】 図4は例えば特開平7-222248号公報に示された従来の音声認識システムの構成を示すブロック図であり、図において、1は端末、2はホスト、3はネットワークである。端末1において、11は表示手段、12はペン入力等の入力手段、13は音声入力手段、14は音声出力手段、15は情報記憶手段、16は通信手段、17はCPUによる演算処理手段、171は演算処理手段17で実行されるプログラムである。

【0003】 一方、図4のホスト2において、21は表示手段、22はキーボード等の入力手段、23は音声入力手段、24は音声出力手段、25は情報記憶手段、26は通信手段、27はCPUによる演算処理手段、28は音声認識手段、271は演算処理手段27で実行されるプログラムである。端末1の通信手段16、ネットワーク2及びホスト2の通信手段26により通信機能が構成される。

【0004】 次に動作について説明する。端末1では、音声入力手段13が音声を入力し符号化して情報記憶手段15に記録する。記録された音声情報は、通信手段16によりネットワーク3を介してホスト2に送信される。ホスト2では、通信手段26が端末1より送信された音声情報を受信し、音声認識手段28が受信した音声情報を認識して文字情報に変換する。通信手段26はその文字情報を端末1に送信する。端末1上に実現される機能及びホスト2上に実現される機能は、各々の内部に

設けられた演算処理手段17及び演算処理手段27が実行するプログラム171及び271で実現されている。

【0005】このようなクライアント・ホスト型のシステム構成において、音声認識のための処理は、ネットワーク3上のいずれかに設置されている、ハードウェア規模が十分に大きなホスト2で行われるので、端末1のハードウェア規模に関係なく、計算処理量が多い高度な音声認識手法を利用することが可能である。

【0006】

【発明が解決しようとする課題】従来の音声認識システムは以上のように構成されているので、生の音声信号の情報量は大きく、この音声信号を符号化してネットワーク3に送信すると、ネットワーク3の負荷が増大するという課題があった。

【0007】また、音声信号を圧縮符号化することで情報量を減らすことも可能であるが、情報量の圧縮に伴う復号化後の音声に品質劣化があるため、圧縮前の音声信号の音声認識の精度に比べて、復号後の音声信号の音声認識の精度が大幅に低下するという課題があった。

【0008】さらに、上記ネットワーク3の負荷増加の課題に対して、特開平3-132797号公報では、端末で、入力音声から音素又は音節を抽出し、これらの記号列を中央処理装置に送信し、中央処理装置で、送信された記号列から単語辞書を使用して入力音声を示す単語や文を認識している。この記号列の伝送量は相当少ないために、ネットワーク3の負荷は小さくなると共に、圧縮による精度低下は解消できるが、入力音声から抽出された音素又は音節には、認識時の誤りが含まれるために、入力音声が文章等の大語彙からなる音声の場合には、一般的な単語辞書を使用した認識結果に十分な精度が得られないという課題があった。

【0009】この発明は上記のような課題を解決するためになされたもので、ネットワーク3の負荷が小さく、文章等の大語彙からなる入力音声の場合にも、認識精度の高い音声認識システムを得ることを目的とする。

【0010】

【課題を解決するための手段】この発明に係る音声認識システムは、ネットワークを介して接続された端末とホストからなるものにおいて、上記端末が、入力した音声信号から、類似の状態が共有化された音素HMMを記憶した音響モデルを使用し、基本単位の列を抽出して上記ホストに送信し、上記ホストが、基本単位の列が示す誤認識の確率を記憶した差分モデルと、単語の連鎖の出現確率を記憶した言語モデルを使用して、受信した上記基本単位の列に整合する単語列を求め、そのテキストを上記端末に送信し、上記端末が受信したテキストを表示するものである。

【0011】この発明に係る音声認識システムは、端末がユーザからの指示に基づき表示しているテキストを修正するものである。

【0012】この発明に係る音声認識システムは、端末が、表示しているテキストを他の端末又は他のホストに送信するものである。

【0013】この発明に係る音声認識システムは、基本単位として音素を使用するものである。

【0014】この発明に係る音声認識システムは、基本単位として音節を使用するものである。

【0015】この発明に係る音声認識システムは、基本単位として音素HMMの並びで表現されるサブワードを使用するものである。

【0016】この発明に係る音声認識システムは、端末が、ユーザからの音声を入力しデジタル化された音声信号を出力する音声入力手段と、上記音声信号から、類似の状態が共有化された音素HMMを記憶した音響モデルを使用して、基本単位の列を抽出する音声認識手段と、抽出された上記基本単位の列をホストに送信すると共に、上記ホストから送信されたテキストを受信する通信手段と、受信した上記テキストを表示する表示手段とを備えたものである。

【0017】この発明に係る音声認識システムは、音声認識手段が、音声信号を入力して、特徴パラメータ時系列と短時間パワー時系列を抽出する音声分析手段と、上記短時間パワー時系列に基づき、音声パワーの存在する可能性の高い音声区間を検出して、上記音声区間内の特徴パラメータ時系列を出力する音声区間検出手段と、類似の状態が共有化された音素HMMを記憶した音響モデルと、この音響モデルに記憶されている音素HMMの組合せの中から、上記音声区間検出手段が出力した特徴パラメータ時系列に整合する基本単位の列を抽出するビタビデコーダとを備えたものである。

【0018】この発明に係る音声認識システムは、ホストが、端末から送信された基本単位の列を受信すると共に、求めたテキストを上記端末に送信する通信手段と、基本単位の列が示す誤認識の確率を記憶した差分モデルと、単語の連鎖の出現確率を記憶した言語モデルを使用して、受信した上記基本単位の列に整合する単語列を求め、テキストとして出力するテキスト変換手段とを備えたものである。

【0019】この発明に係る音声認識システムは、テキスト変換手段が、基本単位の列が示す誤認識の確率を記憶した差分モデルと、テキストを構成する単語の連鎖の出現確率を記憶した言語モデルと、受信した基本単位の列を入力し、上記差分モデルが記憶した基本単位の列が示す誤認識の確率と、上記言語モデルが記憶したテキストを構成した単語の連鎖の出現確率に基づき、上記基本単位の列に整合する単語列を求め、テキストとして出力するスタックデコーダとを備えたものである。

【0020】

【発明の実施の形態】以下、この発明の実施の一形態を説明する。

実施の形態1. 図1はこの発明の実施の形態1による音声認識システムの構成を示すブロック図である。図において、18は端末1において、音声入力手段13からの文章等の大語彙からなる音声を認識して、音節を基本単位とした音節列を求める音声認識手段で、29はホスト2において、通信手段26で受信した端末1からの音節列を入力しテキストを求めるテキスト変換手段であり、その他の構成は、従来の図4における構成と同等のものである。

【0021】図2は端末1における音声認識手段18の構成を示すブロック図である。図において、181は音声入力手段13によりデジタル化された音声信号、182は音声信号181を分析し、特徴パラメータ時系列並びに短時間パワーを抽出する音声分析手段、183は音声分析手段182の出力から、そのうちの短時間パワーに基づいて音声パワーの存在する時間区間を検出し、当該音声区間の特徴パラメータ時系列184を出力する音声区間検出手段、184は音声区間検出手段183によって切り出された音声区間の特徴パラメータ時系列、185は特徴パラメータ時系列184を入力し音声認識結果として、音節を基本単位とした音節列187を出力するビタビデコード、186はビタビデコード185で用いられる音響モデルである。

【0022】図3はホスト2におけるテキスト変換手段29の構成を示すブロック図である。図において、291は通信手段26が受信した音節列、292は音節列291を入力しテキスト295を出力するスタックデコード、293はスタックデコード292が参照する差分モデル、294はスタックデコード292が参照する言語モデルである。

【0023】次に動作について説明する。まず、ユーザからの文章等の大語彙からなる音声は端末1の音声入力手段13に入力されると、音声入力手段13によりデジタル化された音声信号が出力され音声認識手段18に送られる。音声認識手段18は、後述するように音声信号を分析し音節を基本単位とした音節列187を抽出する。音声認識手段18により抽出された音節列187は、通信手段16により、ネットワーク3を介してホスト2の通信手段26に送信される。

【0024】端末1から送信された音節列187は、ホスト2の通信手段26で受信されてテキスト変換手段29に送られる。テキスト変換手段29は、この音節列187を入力し、後述のようにして、端末1で入力した音声信号に対するテキスト295を出力する。通信手段26は、テキスト変換手段29が出力したテキスト295をネットワーク3に送信する。

【0025】端末1の通信手段16は、ネットワーク3からテキスト295を受信する。受信されたテキスト295は表示手段11に表示され、ユーザは表示されたテキスト295により音声認識結果を確認する。表示され

ているテキスト295に誤りがあれば、ユーザからの指示を入力手段12が入力し、演算処理手段17により修正処理が行われる。ユーザは、送信されたテキスト295又は修正されたテキスト295を、通信手段16によりネットワーク3を経由して、図示されていない他のユーザの端末1や他のホスト2に送信する。

【0026】次に図2を用いて音声認識手段18の動作について説明する。音声入力手段13から出力された音声信号181は、音声分析手段182に入力される。音声分析手段182は、音声信号181を入力して、例えば10ms毎に特徴パラメータとしてメルケプストラム係数を抽出する。また、同時に10ms毎に音声区間検出のため短時間パワー時系列を抽出する。これらの時系列は音声区間検出手段183に出力される。ここでは、特徴パラメータとしてメルケプストラム係数を抽出しているが、メルケプストラム係数とデルタメルケプストラム係数の組合せを抽出しても良い。

【0027】音声区間検出手段183は、短時間パワー時系列に基づいて、無音区間や非音声区間を除き、音声パワーの存在する可能性の高い音声区間を検出し、この検出された音声区間内の特徴パラメータの時系列184を出力する。音響モデル186は、前後の音素環境に依存する環境依存型の音素HMM (Hidden Markov Model) からなる。音素HMMは日本語の全音素をカバーするように設計されているが、決定木により類似の状態を共有するような構造となっている。このように、音響モデル186に共有化された音素HMMを使用したことにより、所要のメモリは比較的小さく、端末1のハードウェア規模が小さくて済む。

【0028】この音響モデル186について説明する。音素はleft-to-rightに並んだ1~3個の状態からなっており、各状態は混合ガウス分布で示される出力確率を持っている。テストデータに対して、混合ガウス分布の出力確率が類似している状態は、共通の混合ガウス分布を持たせることにより共有化している。各状態のどれとどれを共有化するかは、決定木を使用したクラスタリングにより決定する。この決定木を使用する方法は、決定木上に置かれた質問に従い、大まかな音素環境から出発し、詳細な音素環境にいたるまで音素環境を変化させることにより、決定木の根の箇所における最大の共有化から出発し、木の枝をたどることにより共有化を分割していくものである。

【0029】ビタビデコード185は、音響モデル186内に記憶された音素HMMのあらゆる組み合わせの中から、特徴パラメータ時系列184に最も良く整合する音素列187を抽出する。これは動的計画法の一種であるビタビアルゴリズムで効率的に実現されている。このようにして、ビタビアルゴリズムで得られた最良の音素列は、音節を基本単位とした音節列187に変換されて出力される。この音素列から音節列の変換は、音素の記

号列の中で母音(a i u e o)、撥音(N)、促音(Q)に着目して、例えば、「watasiwaabetomoosimasu」という音素列を、母音、撥音、促音を1つだけ含む音節列「wa/t a/s i/w a/a/be/to/mo/o/s i/ma/su」という音節列に区切るものである。

【0030】また、ビタビデコーダ185は、ビームサーチ手法を用いて計算回数を削減しているため、演算量が比較的少なく済んでいる。音声区間の始端から終端にかけて行う各フレームのビタビ演算では、始端から当該フレームまでに選択された経路に沿って蓄積された累積尤度が最も高くなるように状態遷移を選択するが、このビームサーチ手法では、各フレームのビタビ演算の際に、累積尤度が閾値以下である経路を枝刈りすることにより経路演算を削減している。

【0031】このように、端末1での音声認識処理は、音節を基本単位とした音節列の認識処理で、言語的な処理を含まないために、すなわち、連鎖の統計量を使用し、ある音節の次に現れる確率の高い音節を求めていく処理を含まないために、演算量が少なく、高速のCPUによる演算処理手段17を備えた端末1で十分高速な処理が可能である。

【0032】次に図3を用いてテキスト変換手段の動作を説明する。差分モデル293は音節列が示す誤認識の確率を記憶したモデルである。誤認識のパターンとして、音節の置換、挿入、脱落があり、音節の置換、挿入、脱落の確率が差分モデル293内に記憶されている。これらの確率は、音声データベースに集められた音声認識して得られる音声認識結果の基本単位の列(音節列)と、音声データベースに付けられている正解の基本単位の列(音節列)を比べて、正解に対する認識結果における基本単位(音節)の置換、挿入、脱落をカウントすることにより推定する。

【0033】言語モデル294は、テキストを構成する単語の連鎖の出現確率が記憶されている。単語の連鎖としてはn単語の連鎖を用い、当該単語連鎖の出現確率として記憶されている。この単語連鎖の出現確率は、対象と類似した文例を多数集めたコーパスから推定される。このコーパスは、電子化されたべた書きの文例に対して、単語連鎖の統計量を求めるために、予め単語間を分割して記憶したものである。この実施の形態では、nは1から3の範囲の値を取る。

【0034】この実施の形態における言語モデル294は、新聞記事等の一般のテキストを対象としているため、数百Mバイトの記憶容量が必要であるが、ホスト2上に設置していることにより、言語モデル294を全てメモリ上に展開して処理してもメモリ上の問題は生じない。

【0035】また、演算処理手段27が演算能力の高いCPUを用いているので、処理時間は実用上十分な高速

処理が可能であり、音声認識処理システムとして応答性の問題も生じない。スタックデコーダ292は、音節列291を入力して、差分モデル293に記憶された音節の誤認識の確率と、言語モデル294に記憶された単語連鎖の確率を加味して、音節列291に最も良く整合する単語列を求め、これをテキスト295として出力する。

【0036】なお、この実施の形態では、音節を基本単位としているが、音素を基本単位としても良い。この場合、端末1から、ビタビデコーダ185がビタビアルゴリズムで求めた最良の音素列、例えば、「watasiwaabetomoosimasu」という音素列をそのまま送信し、ホスト2では、スタックデコーダ292が、受信した音素列を入力し、差分モデル293に記憶された音素の誤認識の確率と、言語モデル294に記憶された単語連鎖の確率を加味して、受信した音素列に最も良く整合する単語列を求め、これをテキスト295として出力する。

【0037】また、この実施の形態では、音素や音節の代わりに、音素の並びで表現されるサブワードを基本単位としても良い。この場合、端末1から、ビタビデコーダ185がビタビアルゴリズムで求めた最良の音素列、例えば、「watasiwaabetomoosimasu」という音素列を、「watasiwa/abeto/moosimasu」というサブワード列に変換して送信し、ホスト2では、スタックデコーダ292が、受信したサブワード列を入力し、差分モデル293に記憶されたサブワードの誤認識の確率と、言語モデル294に記憶された単語連鎖の確率を加味して、受信したサブワード列に最も良く整合する単語列を求め、これをテキスト295として出力する。

【0038】基本単位として、音素、音節、サブワードを使用した場合を比較すると、差分モデル293に基本単位の置換、挿入、脱落の確率が記憶されているときは、長い単位の方が、すなわち、音素より音節、音節よりサブワードの方が、広範囲の文脈を考慮した認識誤りに対処できる。

【0039】以上のように、この実施の形態1によれば、端末1の音声認識手段18が、小規模な音響モデル186を参照して、基本単位の列(音素列、音節列、又はサブワード列)を求めて、ネットワーク3を介してホスト2に送信し、ホスト2では、受信した基本単位の列を、差分モデル293と大きな記憶容量を持つ精度の高い言語モデル294を参照して、テキスト295に変換するようにしているので、端末1とホスト2間のデータ転送量が小さいためにネットワーク3の負荷が小さく、ハードウェア規模の小さい端末1に入力された音声、文章等の大語彙からなる音声の場合にも、高速かつ認識精度の高い音声認識システムを得ることができるという効果が得られる。



【0040】また、この実施の形態1によれば、端末がユーザからの指示に基づき表示しているテキストを修正することにより、ホストから送信された誤認識の結果を訂正することができると共に、端末が、表示しているテキストを他の端末又は他のホストに送信することにより、ホストからの認識結果を他の端末又は他のホストに通知でき、認識結果を共有化できるという効果が得られる。

【0041】さらに、この実施の形態1によれば、基本単位として音素、音節、又はサブワードを使用することにより、ネットワークの負荷を小さくできると共に、長い単位を使用した場合には、広範囲の文脈を考慮した認識誤りに対処できるという効果が得られる。

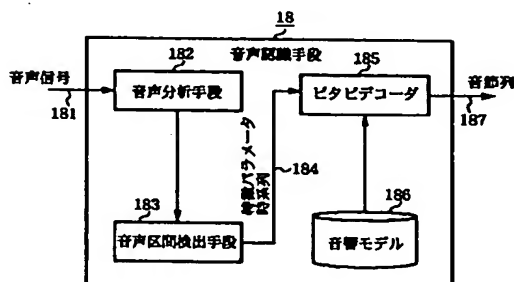
【0042】

【発明の効果】以上のように、この発明によれば、端末が、入力した音声信号から、類似の状態が共有化された音素HMMを記憶した音響モデルを使用し、基本単位の列を抽出してホストに送信し、ホストが、基本単位の列が示す誤認識の確率を記憶した差分モデルと、単語の連鎖の出現確率を記憶した言語モデルを使用して、受信した基本単位の列に整合する単語列を求め、そのテキストを端末に送信し、端末が受信したテキストを表示することにより、ネットワークの負荷が小さく、端末に入力された音声、文章等の大語彙からなる音声の場合にも、高速かつ認識精度の高い音声認識システムを得ることができるという効果がある。

【0043】この発明によれば、端末がユーザからの指示に基づき表示しているテキストを修正することにより、ホストから送信された誤認識の結果を訂正することができるという効果がある。

【0044】この発明によれば、端末が、表示しているテキストを他の端末又は他のホストに送信することにより、ホストからの認識結果を他の端末又は他のホストに通知でき、認識結果を共有化できるという効果がある。

【図2】



【0045】この発明によれば、基本単位として音素を使用することにより、ネットワークの負荷を小さくできるという効果がある。

【0046】この発明によれば、基本単位として音節を使用することにより、ネットワークの負荷を小さくできるという効果がある。

【0047】この発明によれば、基本単位として音素HMMの並びで表現されるサブワードを使用することにより、ネットワークの負荷を小さくできると共に、広範囲の文脈を考慮した認識誤りに対処できるという効果がある。

【図面の簡単な説明】

【図1】 この発明の実施の形態1による音声認識システムの構成を示すブロック図である。

【図2】 この発明の実施の形態1による音声認識手段の構成を示すブロック図である。

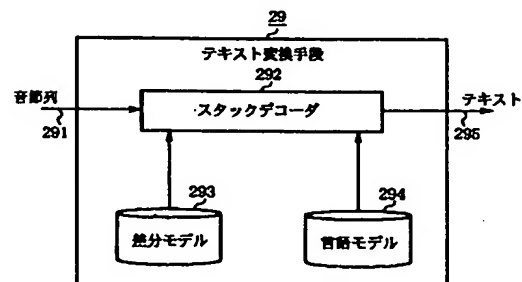
【図3】 この発明の実施の形態1によるテキスト変換手段の構成を示すブロック図である。

【図4】 従来の音声認識システムの構成を示すブロック図である。

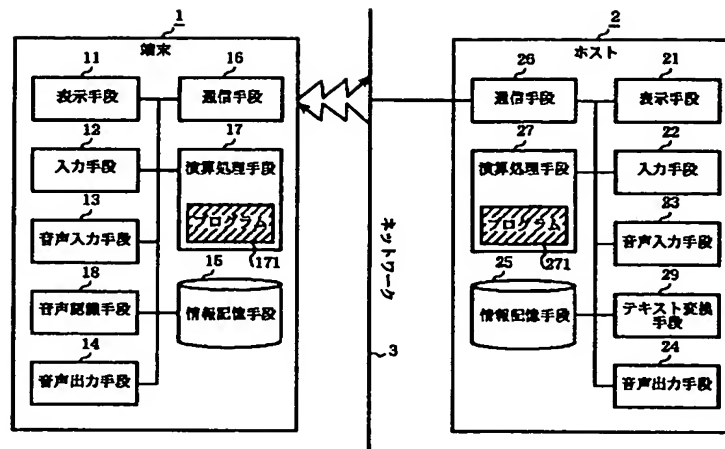
【符号の説明】

1 端末、2 ホスト、3 ネットワーク、11 表示手段、12 入力手段、13 音声入力手段、14 音声出力手段、15 情報記憶手段、16 通信手段、17 演算処理手段、18 音声認識手段、21 表示手段、22 入力手段、23 音声入力手段、24 音声出力手段、25 情報記憶手段、26 通信手段、27 演算処理手段、29 テキスト変換手段、171 プログラム、271 プログラム、181 音声信号、182 音声分析手段、183 音声区間検出手段、184 特徴パラメータ時系列、185 ビタビデコーダ、186 音響モデル、187 音節列、291 音節列、292 スタックデコーダ、293 差分モデル、294 言語モデル、295 テキスト。

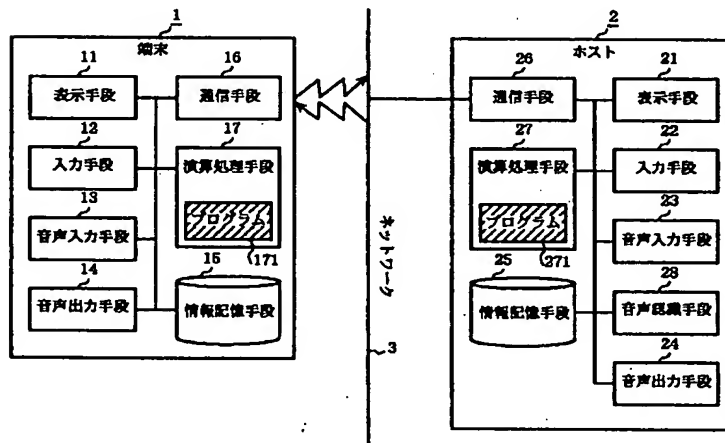
【図3】



【図1】



【図4】



フロントページの続き

(51)Int.Cl.<sup>7</sup>  
G10L 15/22

識別記号

FI  
G10L 3/00

テームド (参考)

561E  
561F